https://doi.org/10.2352/ISSN.2470-1173.2017.16.CVAS-343

Aerodynamic analysis via foreground segmentation

Peter Carey[1], Stuart Bennett[1], Joan Lasenby[1], and Tony Purnell[1,2]
[1]Cambridge University Engineering Department, Cambridge, UK
[2]British Cycling, UK

Abstract

Results from wind-tunnel testing of athletes cannot always be repeated on the track, but reducing aerodynamic drag is critical for racing. Drag force is highly correlated with an athlete's frontal area, so in this paper we describe a system to segment an athlete from the very challenging background found in a standard racing environment. Given an accurate segmentation, a front-on view, and the athlete's position (for scaling), one can effectively count the pixels and thereby measure the moving area. The method described does not rely on alteration of the track lighting, background, or athlete's appearance. An image-matting algorithm more used in the film industry is combined with an innovative model-based pre-process to allow the whole measurement to be automated. Area results have better than one percent error compared to hand-extracted measurements over a representative period, while frame-by-frame measurements capture expected cyclic variation. A near real-time implementation permits rapid iteration of aerodynamic experiments during training.

Introduction

In racing sports, aerodynamics are as important as power, and small improvements affect medal places. Wind-tunnel testing is useful but expensive, and having a stationary athlete may lead to pose optimizations that cannot be reproduced or maintained in reality. In this work we quantify aspects of aerodynamic performance in the training/competition environment of a velodrome.

Aerodynamic drag is proportional to a body's frontal area. We therefore wish to accurately segment the athlete from the background as they move towards the camera, and so find the apparent frontal area. The two main problems are distinguishing the athlete from the background and excluding the moving athlete's shadow from the measurement. Compounding the difficulty is having control over none of the background (no green screen), the rider's clothing/appearance (skin similar in colour to velodrome wood), or the lighting — see Figure 1 below, where the skin and wood colours lack contrast, and the black tyre seems to merge into the line marked on the track.

The application of matting to the topic of real-time sports analysis, rather than its customary home of film and animation, especially in the context of quantitative aerodynamic performance assessment, is novel. The image-matting algorithm itself is heavily based on 'Fast matting using large kernel matting Laplacian matrices' by He, Sun and Tang [1]. There is relatively little published on fast automatic generation of trimaps for video input, with most literature assuming the trimap is a user-supplied input.

In the next section we describe the design and implementation of the measuring process in detail, starting from an unanalysed input video. Subsequent sections show sample analysed output, and discuss the performance of the system as a whole.

Figure 1. *Difficult to distinguish: leg vs. track, tyre vs. black line*

System description

Video is recorded by a sensitive 50 f.p.s. camera, located on-axis with the velodrome straight, and triggered as the cyclist enters the straight. Automatically adjusting focus (or indeed zoom) in real-time is impractical, as the bicycles are moving rapidly toward the camera, and there may be more than one bicycle on the straight at a time. Therefore we instead set the camera to have a small aperture for a high depth of field, and a correspondingly high sensor gain.

A three-stage image-processing pipeline is used. First, the bicycle is located through motion estimation and adaptive colour-thresholding. Second, in the region of the detected bicycle several frames are combined to form an estimate of the background, and a 'trimap' is automatically generated, representing areas known to be foreground, background, or uncertain, informed by the expected shape of the bicycle/cyclist combination. Finally, we apply a

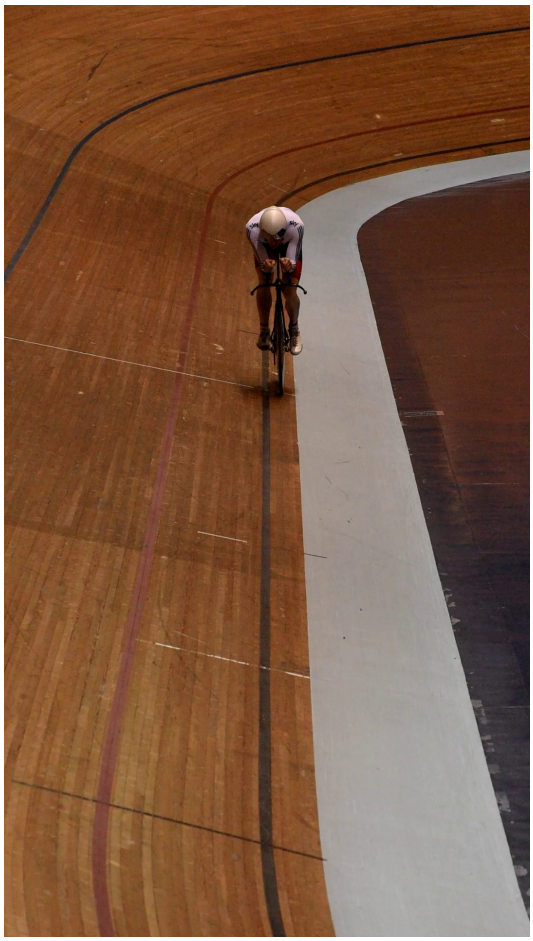

Figure 2. *Source frame: height of contact point implies distance to cyclist and expected cyclist size. Note shadows around tyre and similarity of colour of clothing and light-blue background.*

natural image-matting algorithm to estimate the true nature of the uncertain areas based on colour statistics of spatially nearby regions of known fore-/back-ground. The output yields a map of each pixel's probability of belonging to the foreground; after thresholding, summing, and scaling, the frontal area is found. These stages are described in turn below.

Bicycle discovery

Locating the bicycle within the image frame fundamentally relies on a bicycle's front wheel's tyre being black – locally the darkest object on the track. Additional techniques are necessary due to similarity to the track's black line, and shadows cast by overhead lighting.

The first process is to generate a map of areas exhibiting change (i.e. motion) by subtracting the luminance ('Y', in a YUV colour model) channel of the current frame from that of the frame 100 ms previously, downsampling and binarizing that image, performing morphological opening, small object removal, and dilation operations ([2]), and upscaling the result to the size of the original frame. The result of this is to suppress minor changes and grow the boundaries of regions of gross change.

Second, the height up the frame of the lowest contact point is sought (though starting from half-way up the frame early in the clip to avoid accidental detection of any bicycle preceding the one triggering the camera's recording). We begin by thresholding the luminance channel for dark objects and intersecting the result with the previously gained motion map. Again, small dark objects are discounted, first by downsampling, then by morphological small object removal operations adapted to the 'height' of the results from the downsampling step (bicycles 'higher' in the frame are further away and hence smaller, as depicted in Figure 2). If, at this point, no dark objects remain, due to the bicycle triggering the recording having moved out of frame, no height is recorded, otherwise processing continues. Next we retain only those objects nearest the camera which have some vertical range (i.e. look like an upright tyre). These objects (more than one occurring during an overtake manoeuvre) proceed to 'fine' height estimation, with the lowest/nearest being adopted as the contact point of the nearest bicycle. Fine height estimation compares the area around the coarsely obtained height on the original luminance image with a threshold based on the darkest intensity present in the downsampled image of the same area, and again applies the motion map as a mask. Any small objects ('small' again defined by the height/nearness of the bicycle) are culled and the mean height of the lowest/nearest ten dark pixels is taken as the contact point, while the coarse lateral position is also recorded for later refinement.

For robustness, before precisely finding the lateral positions of the contact points, the set of heights from the whole recording is now analysed for continuity. Bicycle motion is very smooth, generally with little change in velocity, so we can detect frames where another bicycle tyre, or something other than a tyre, was returned as the primary contact point and interpolate these heights (and coarse lateral positions) by fitting smooth splines to the accepted data. The details of discounting frames from a multiple bicycle situation are not detailed here, as the area extraction results will not have useful meaning in such cases. For detection of single-frame glitches an approach such as LOWESS (locally weighted scatterplot smoothing, from [3]) is adequate.

Refining the lateral positions is a process similar to that for fine height estimation. Again the motion-containing darker-than-all-else pixels of the area around the coarse estimate are considered. Knowing the distance (height) to the wheel, the expected pixel width of the tyre is known. A window of this width, and roughly half a tyre high, is passed over all plausible lateral positions, and a score formed based on the number of dark pixels found in the window (pixels in the lower half of the window count double, as the tyre may be inclined from vertical): the centroid of the dark pixels for the window position with the highest score is taken as the lateral contact co-ordinate. The set of lateral co-ordinates then undergoes a LOWESS outlier discarding and spline smoothing and interpolation process.

Background estimation and coarse segmentation

We use the capture video to generate the background model, as no other clip would contain as relevant lighting conditions. Presuming that cyclist segmentation is only worth attempting in conditions where only one cyclist is in shot leads to a simple method of estimating the background. The background model is formed by stacking the median-filtered average of the upper-half of the last five frames of the clip (when the cyclist will be low in

the frame) on top of the median-filtered average of the lower-half of the first five frames (where the cyclist will be entering the frame at the top).

With the set of smoothed contact point co-ordinates known, each frame is cropped to loosely bound the cyclist, with the front wheel contact point at a known offset in the crop. The expected shape and size of the cyclist (deduced from the track position/distance from the camera) is then used to tailor a 'trimap' of the frame, the pixels being labelled as:

1. parts believed to be background,
2. parts believed to be foreground (i.e. the bicycle and cyclist), and
3. the uncertain areas which may be foreground or background.

The better this trimap is the quicker and better the computation of the final segmentation results.

Working in RGB colour space, we begin by thresholding the difference of the green channels from the current and background frames. This gives an approximate segmentation of the moving cyclist, and is refined further using greyscale morphological operations, the first of the which is an opening, to give the basic mask.

The bottom of the front wheel often needs refinement due to the presence of shadows. Working in all three colour channels the blackest part in the region above the contact point is dilated and kept, all other parts of the mask in this area are discarded. With known scale and wheel position, the position in the image of other body- and bicycle-parts may be inferred, such as the cyclist's legs, and the bicycle's stem and saddle. The basic mask often has erroneous holes, due to parts of the cyclist appearing similar to the background, as shown in Figure 1 and Figure 2. A large structuring element is used for a morphological close operation on the mask above the saddle, covering the cyclist's head and torso. Below the saddle a narrow region close to the bicycle frame is closed in two parts: for the part between the stem and saddle, mostly comprising frame, arms and thighs, a medium-sized element is used, while a small element is used below the stem, where the finer details of the lower legs and structural elements of the bicycle are present. As the legs are often wider than this narrow region, and relatively large and solid, a larger element is used in a closing operation either side of the bicycle-frame region.

These mask manipulations complete, the greyscale mask is thresholded, and the largest contiguous area taken as being the cyclist-with-bicycle combination and labelled as foreground (shown as white in Figure 3). The areas not covered by a dilation of the foreground are taken to be background (black), and the areas of difference between foreground and background are labelled as uncertain (grey).

The trimap is loosely cropped to the bounding box of the uncertain areas, meaning that all pixels identified as background are near the cyclist and so have properties more relevant to describing the areas obscured by the cyclist than those background pixels further away.

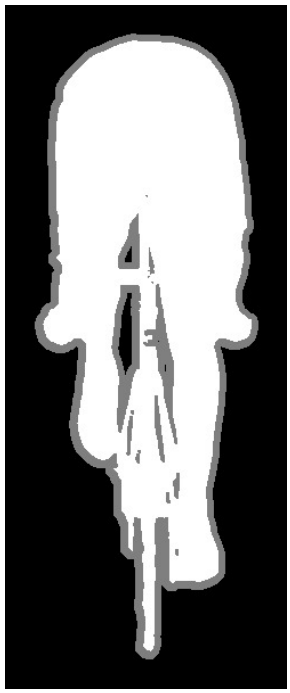

Figure 3. *Loosely cropped trimap: background shown as black, foreground (cyclist) as white, uncertain as grey*

Image matting and area calculation

The cropped trimap and corresponding region of the current frame are passed to the image-matting algorithm. We employ an efficient implementation of He et al.'s 'Fast matting using large kernel matting Laplacian matrices' algorithm [1], which uses integral images ([4]) to speed up computation. The paper's proposed KD-tree pre-process is not performed, as we have found that, in our application, it doesn't notably improve segmentation results, but does add to the computation time.

In brief, the model matting approaches adopt is one where a pixel's colour $\mathbf{I}$ is a combination of the foreground ($\mathbf{F}$) and background ($\mathbf{B}$) colours, moderated by the foreground opacity α:

$$\mathbf{I} = \mathbf{F}\alpha + \mathbf{B}(1-\alpha). \tag{1}$$

Pixels where α is 0 or 1 therefore belong to the background or foreground respectively. Since our input is $\mathbf{I}$, the challenge is to estimate α everywhere in the uncertain region of the trimap. Contemporary matting approaches exploit a *colour line* assumption: that in RGB space the foreground (or background) colours in a spatial neighbourhood lie along a single line. From this a cost function function can be formed to determine the alpha value best relating a candidate pixel's colour to the known nearby foreground and background colours, and this cost function optimized. While the approach can be hindered by complicated foreground/background colours, we find that in our application the correct formation of the trimap is more critical to the accuracy of the results.

Having supplied or estimated α for all the pixels in the frame, the area of the cyclist-with-bicycle combination can simply be found by counting the number of pixels mostly belonging to the foreground, i.e. $\alpha > 0.5$. This area is then scaled to counteract the cyclist appearing to enlarge as they approach the camera.

Results

The bicycle-location aspect has been used on over 90,000 clips, with sampled accuracy over 98 %, i.e. with the triggering bicycle's position correctly and smoothly tracked.

Sample segmentation output can be seen in Figure 4, set against a green background for contrast. While largely correct, some small errors can be seen, with a hole in the cyclist's left thigh in the leftmost frame, partial inclusion of some shadows and track texture in the rightmost frame, and omission of the area containing the bicycle-frame manufacturer's logo from the head tube in all three frames.

In general, however, the snapshot nature of single frames makes consideration of their errors less useful; for quantitative output we prefer to average over a pedal revolution. Compared to hand-extracted frontal-area data the automatic segmentation has less than one percent error over a revolution.

An example where pose changes affect the longer-term averages is shown in Figure 5. Each time the pedals are level (twice per complete revolution) is shown by a red vertical line, and we see that *per frame* the area is somewhat noisy, while the overall shape is clear: a gradual expansion in area, followed by a sharp contraction. This occurred when a cyclist allowed their elbows to drift apart, followed by a correction to a much tighter pose as they realized their lapse.

Within-revolution data are nonetheless valuable for spotting certain behaviours more qualitatively. In Figure 6, the cyclic variation in area throughout the revolution is highly apparent: bearing in mind the red lines indicate a *half* revolution, the pronounced increase in area *once* per revolution suggest an asymmetry, presuming the cyclist approaches the camera on-axis (while largely obscured by the gross changes, the expected half-revolution peaks can be discerned in Figure 5).

The apparent gradual reduction in area over time observed in both Figure 5 and Figure 6 is due to the camera being installed above the track, and perspective effects causing later frames to be seen increasingly from above, rather than purely front-on. A simple geometrical correction for this may be applied as a post-process.

Implementation work has made the processing of a four second clip, covering the whole straight, take under twenty seconds, meaning the results are available in near real-time.

Conclusions

In this paper we have described a system for accurately locating a bicycle in a velodrome environment, automatically segmenting it from the background, and measuring the apparent area presented by the bicycle and cyclist. The system works consistently on challenging real-world data in an uncontrolled environment, where track markings, lighting, shadows and multiple bicycles may impact accurate location, and similarity of colour of cyclist skin and clothing may be hard to distinguish against the background.

The resultant area measurements are useful for assessing aerodynamic performance, a critical component of modern track racing, while doing so in a competition environment, rather than in artificial wind-tunnel conditions. Overall such an approach could have potential in other sports, or, more widely, other fields needing automated segmentation.

Figure 4. *Selection of segmented and identically scaled frames, with background pixels set to green*

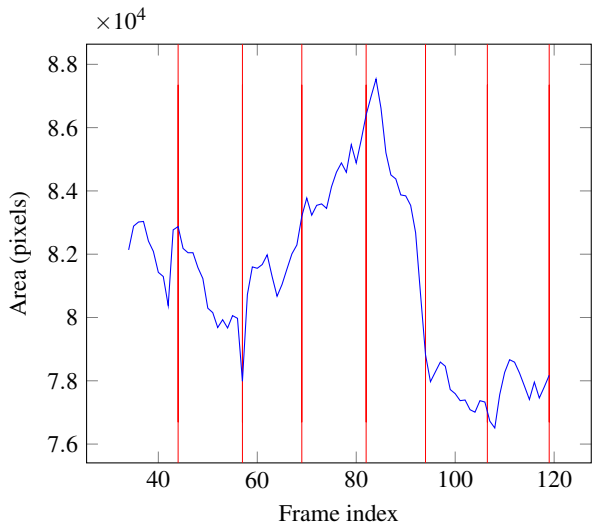

Figure 5. *Example area measurements (blue line), vertical bars indicate the 'feet level' condition — half a pedal revolution. A gradual expansion in area, followed by a sharp contraction.*

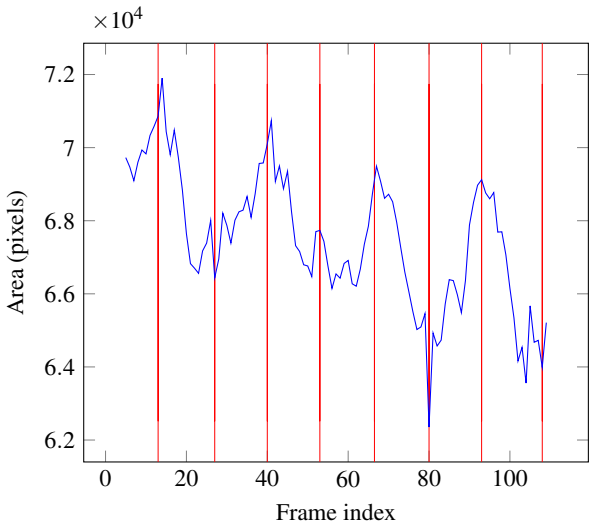

Figure 6. *Example area measurements (blue line), vertical bars indicate the 'feet level' condition — half a pedal revolution. Cyclic area variation suggests an asymmetry.*

References

[1] Kaiming He, Jian Sun, and Xiaoou Tang. Fast matting using large kernel matting Laplacian matrices. In *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR 2010, pages 2165–2172. IEEE Computer Society, June 2010.

[2] Robert M. Haralick, Stanley R. Sternberg, and Xinhua Zhuang. Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):532–550, April 1987.

[3] William S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, December 1979.

[4] Franklin C. Crow. Summed-area tables for texture mapping. In *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '84, pages 207–212. ACM, July 1984.

Author Biography

Peter Carey graduated from the University of Cambridge MEng course in 2016. During his undergraduate career he undertook two placements with the Department of Engineering's Undergraduate Research Opportunities Programme, working on implementing and evaluating image-matting algorithms. His master's dissertation researches the development of segmentation algorithms specialized for a velodrome environment.

Stuart Bennett gained his MEng in Information Engineering in 2007, and later his PhD on real-world 3D reconstruction in 2014, from the University of Cambridge, supervised by Joan Lasenby, in the Signal Processing Group of the Department of Engineering. Now a postdoctoral researcher, his research interests span computer vision, with a fondness for real-time robust image processing and multiple-view reconstruction applications with implementation challenges.

Joan Lasenby studied Mathematics at Cambridge University, then spent a year as a TA in Louisiana State University. Her subsequent PhD with the Radio Astronomy group in the Department of Physics, Cambridge, was followed by a Junior Research Fellowship at Trinity Hall, Cambridge, and working for Marconi Maritime Underwater Systems. Returning to academia, firstly as a postdoctoral researcher, then as a Royal Society University Research Fellow, she is currently a Reader in the Signal Processing Group of the Cambridge University Engineering Department. Her research involves image processing, motion capture, human motion modelling, medical applications of vision and geometric algebra.

Tony Purnell studied Engineering at Manchester University, before moving to MIT as a Kennedy Scholar, and then completing his PhD as a Benefactor's Scholar of St John's College, Cambridge. In 1986 he founded Pi Research, developing automotive analysis and control systems, sold to Ford in 1999. In the following years he has been principle of the Jaguar/Red Bull Formula One racing teams, and a technical advisor to the FIA. A Royal Academy of Engineering visiting professor at the University of Cambridge, in 2013 he became head of technical development at British Cycling.